

Association rule/pattern mining for recommender system

**Group**:

Ritika Giridhar a1907667

Sakshi Sinha a1898508

**The University of Adelaide**

4433\_COMP\_SCI\_7306 Mining Big Data  
 Lecturer: Dr. Alfred Krzywicki

**Table of Contents**

[1. Executive Summary 4](#_Toc163725314)

[2. Introduction 4](#_Toc163725315)

[2.1. Collaborative filtering 4](#_Toc163725316)

[2.1.1. Cosine similarity 5](#_Toc163725317)

[2.2. Emergence of frequent pattern mining 5](#_Toc163725318)

[2.3. Association rule mining 5](#_Toc163725319)

[2.4. Recommendation system overview 6](#_Toc163725320)

[3. Exploratory Analysis 6](#_Toc163725321)

[3.1. Data and dataset overview 6](#_Toc163725322)

[3.2. Analysis of quantity and cost distributions 7](#_Toc163725323)

[3.3. Top item pairs 7](#_Toc163725324)

[3.4. Analysis of frequently purchased items 7](#_Toc163725325)

[3.5. Analysis of temporal trends 8](#_Toc163725326)

[3.6. Transaction volume by day of the week 8](#_Toc163725327)

[3.7. Top items revenue contribution 9](#_Toc163725328)

[4. Implementation and Testing 9](#_Toc163725329)

[4.1. Recommendation system design 9](#_Toc163725330)

[4.2. Breakdown of part-wise implementation 10](#_Toc163725331)

[4.2.1. Data preprocessing 10](#_Toc163725332)

[4.2.2. Encoding users and items 10](#_Toc163725333)

[4.2.3. Generating user-item interaction matrix 11](#_Toc163725334)

[4.2.4. Computing item-item similarity 11](#_Toc163725335)

[4.2.5. Identifying frequent item pairs 12](#_Toc163725336)

[4.2.6. Personalising recommendations 12](#_Toc163725337)

[4.2.7. Generating recommendations 12](#_Toc163725338)

[4.3. Methodology 12](#_Toc163725339)

[4.3.1. Training and tuning on the training set 12](#_Toc163725340)

[4.3.2. Using the test set for final results 13](#_Toc163725341)

[4.3.3. Evaluation methodology 13](#_Toc163725342)

[5. Discussion of Results 13](#_Toc163725343)

[5.1. Frequent patterns 13](#_Toc163725344)

[5.1.1. Training set patterns 13](#_Toc163725345)

[5.1.2. Testing set patterns 13](#_Toc163725346)

[5.2. 10 examples of recommendations 14](#_Toc163725347)

[5.3. Discussion of metrics 15](#_Toc163725348)

[5.3.1. Performance with and without pattern mining 15](#_Toc163725349)

[5.3.2. Impact of pattern mining 15](#_Toc163725350)

[6. Conclusion and Recommendations 15](#_Toc163725351)

[6.1. Addressing challenges 16](#_Toc163725352)

[6.2. Recommendations for improvement 16](#_Toc163725353)

[7. Reflection 16](#_Toc163725354)

[8. References 17](#_Toc163725355)

[*Appendix 1* 18](#_Toc163725356)

[Glossary of Terms 18](#_Toc163725357)

**List of Figures**

[Figure 1 Example of cross referencing 4](#_Toc163725201)

[Figure 2 Vectoral representation of cosine similarity 5](#_Toc163725202)

[Figure 3 Frequent pattern generation 5](#_Toc163725203)

[Figure 4 Association rule mining for itemsets 6](#_Toc163725204)

[Figure 5 Recommendation system based on collaborative filtering 6](#_Toc163725205)

[Figure 6 Quantity and price distribution of the dataset 7](#_Toc163725206)

[Figure 7 Top 20 most frequently purchased items 8](#_Toc163725207)

[Figure 8 Monthly transaction volume 8](#_Toc163725208)

[Figure 9 Transaction volume by day of the week 9](#_Toc163725209)

[Figure 10 Top 10 items by revenue contribution 9](#_Toc163725210)

[Figure 11 Overview of the recommendation system 10](#_Toc163725211)

[Figure 12 Overview of data preprocessing 10](#_Toc163725212)

[Figure 13 Overview of user and item encoding 11](#_Toc163725213)

[Figure 14 Overview of user-item matrix generation 11](#_Toc163725214)

[Figure 15 Overview of item-item similarity computation 11](#_Toc163725215)

[Figure 16 Overview of identification of frequent item pairs 12](#_Toc163725216)

[Figure 17 Overview of personalised recommendation generation 12](#_Toc163725217)

[Figure 18 Overview of recommendation generation 12](#_Toc163725218)

**List of Tables**

[Table 1 Frequent item pairs and their count 7](#_Toc163725579)

[Table 2 Top 5 frequent item pairs in the train dataset and their support 13](#_Toc163725580)

[Table 3 Top 10 frequent patterns in the training set and their support 14](#_Toc163725581)

[Table 4 Top 10 frequent patterns in the testing set and their support 14](#_Toc163725582)

# Executive Summary

The development and integration of recommender systems (RS) has become critical for improving customer experiences in various fields, including e-commerce. At our online grocery store, we have embraced this trend and adopted an efficient solution that not only boosts sales, but also increases customer satisfaction. By analysing the user’s transaction data, we discovered frequently co-purchased items, allowing us to intuitively recommend things to online customers (Gambito Analytics, [2023](#_References)). Our integrated recommendation system generates personalised recommendations for consumers using advanced approaches like pattern mining, frequency, and transaction data analysis. By doing so, we not only respond to individual preferences, but also offer value to our online platform (Silva, [2020](#_References)).

Personalised recommendations have the potential to boost conversion rates and income—they can improve client satisfaction, resulting in higher retention rates. Furthermore, controlled retailing allows us to proactively influence purchase behaviour and promote specific goods, thereby improving sales performance. Our approach is an intelligent use of data mining tools to solve real-world business problems. As we continue to modernise the e-commerce space, our focus remains on delivering exceptional experiences that exceed expectations and foster long-term customer loyalty. Thus, we are well-positioned to drive our grocery store’s performance and expansion by employing transaction data to optimise suggestions and improve customer experiences.

# Introduction

The growth of RSs has revolutionised how companies interact with customers. With a large amount of information available online, customers often get overwhelmed, especially in scenarios like online shopping where they are presented with extensive lists of items to choose from. RS address this challenge by leveraging algorithms to predict user preferences or ratings on items, thereby helping users in finding items of interest amidst the overwhelming array of choices.

Traditional RS techniques, such as collaborative filtering, content-based filtering, and hybrid methods, have been around for decades. However, recent advancements in machine learning and data analysis have enabled the development of more sophisticated hybrid methods (Zhang et al., [2020](#_References); Verma & Sharma, [2021](#_References)). These newer approaches leverage the strengths of multiple techniques (Afsar et al., [2022](#_References)) to provide even more accurate and personalised recommendations, further enhancing the user experience and driving business success in today’s competitive landscape.

## Collaborative filtering

Collaborative filtering is a technique filters out items a user might like on the basis of responses by other similar users.

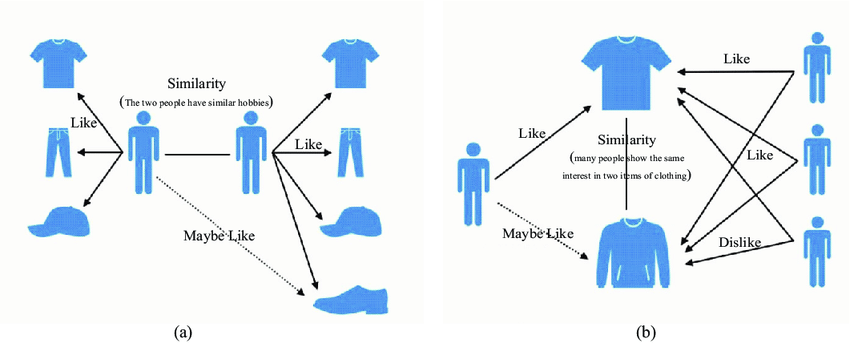


Figure 1 Example of cross referencing (Source: Jianjun et al., [2021](#_References))

It works by searching a large group of people and finding a smaller set of users with tastes similar to a particular user. It looks at the items they like and combines them to create a ranked list of suggestions.

## Cosine similarity

Cosine similarity is another technique used in recommendation systems, particularly in collaborative filtering.

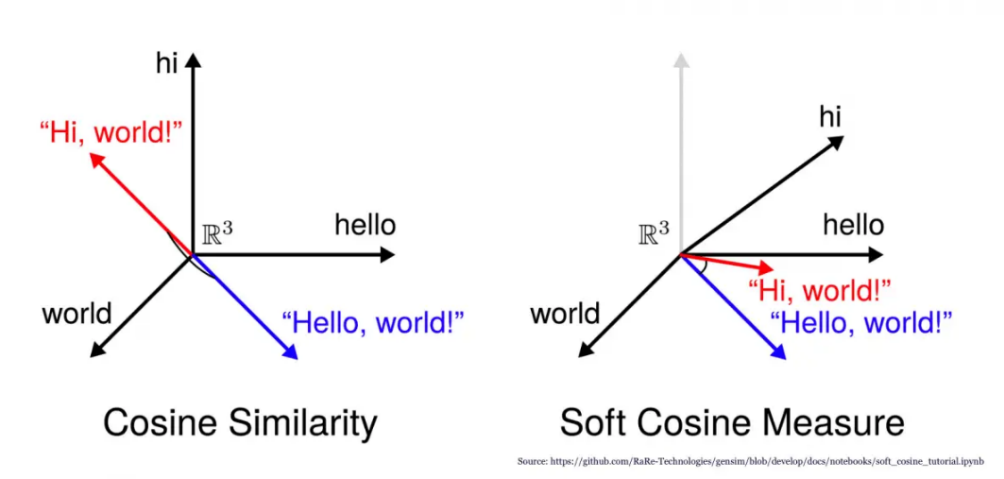


Figure 2 Vectoral representation of cosine similarity (Source: Prabhakaran, [2024](#_References))

It measures the cosine of the angle between two vectors, typically representing user-item interactions or item-item similarities. By calculating the cosine similarity between users or items, recommendation systems can identify similarities and make relevant recommendations based on these similarities.

## Emergence of frequent pattern mining

In addition to traditional RS techniques, frequent pattern mining has emerged as a valuable approach to understanding customer behaviour and enhancing recommendation systems. Frequent patterns are patterns that appear frequently in a dataset, indicating strong correlations among items.

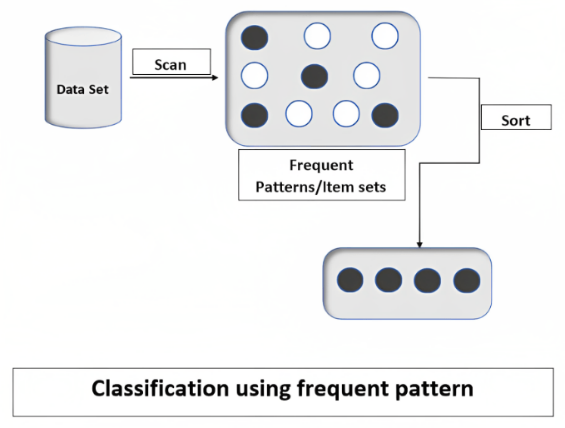


Figure 3 Frequent pattern generation (Source: GeeksforGeeks, [2023](#_References))

By identifying frequent patterns, businesses can observe associations among items, identify similar characteristics, and uncover valuable insights into customer preferences.

## Association rule mining

Association rule mining, a subset of frequent pattern mining, works by analysing large datasets of transactions or events, identifying frequent itemsets, and generating rules describing associations between items. These rules form the basis for cross-selling strategies in the retail industry, suggesting complementary products to customers based on past purchases or interests.

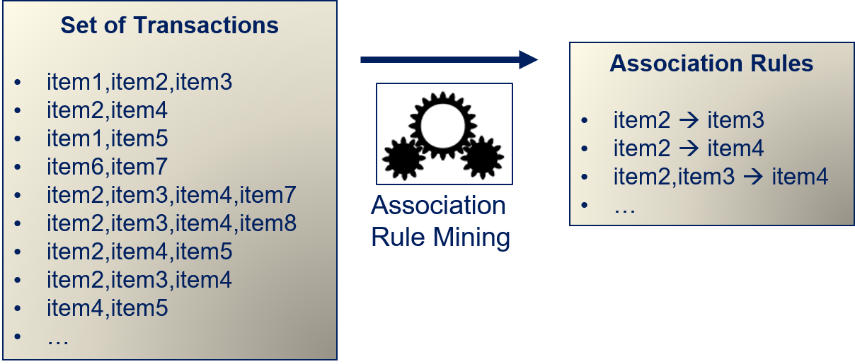


Figure 4 Association rule mining for itemsets (Source: Herbold, [2020](#_References))

Association rule mining can enhance recommendation systems by analysing user ratings, feedback, or behavioural data to identify correlations between items (LinkedIn, [2023a](#_References), [2023b](#_References)). By stating rules that capture these associations, these systems can provide personalised recommendations to users based on their interactions with related items, thereby enhancing user experience and increasing engagement.

## Recommendation system overview

Our recommendation system includes collaborative filtering (based on item and cosine similarity) and association rule mining (using frequent item pairs). The system generates a list of recommended items for each item in the user’s basket, taking into account both direct and cosine similarity, as well as the influence of frequent item pairings.

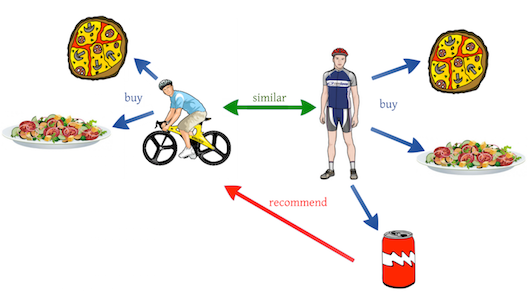


Figure 5 Recommendation system based on collaborative filtering (Source: Kirk, [2019](#_References))

Recommendations are deduplicated and limited to the top N items to maintain relevance and diversity. This approach blends collaborative filtering’s ability to capture user preferences via item similarity and cosine similarity with insights from association rule mining on item co-occurrences (Kryzwicki, [2024](#_References)). As a result, it provides a reliable mechanism for making tailored and contextually relevant recommendations. The addition of a method for unseen objects ensures that the system is adaptable to new inventory or changing user preferences.

# Exploratory Analysis

## Data and dataset overview

The dataset **basket\_data\_by\_date\_train\_big.csv** contains 200,000 entries detailing transactions involving 3,598 distinct *Itemnames*, 12,163 unique *BillNos*, and 3,643 individual *CustomerIDs*. Across these transactions, the *Quantity* of items purchased spans from 1 to 10, with an average purchase quantity of about 3.72 items per transaction. The *cost* of items demonstrates considerable variation, ranging from $0.06 to $295, indicating a broad spectrum of item values. Similarly, the total transaction cost varies significantly, falling between $0.10 and $527.70, with an average transaction cost of approximately $11.23. Additionally, transactions are distributed across all days of the week, with the *weekday\_transactions* variable representing days from 0 for Monday to 6 for Sunday.

## Analysis of quantity and cost distributions

We visualised the distributions of *Quantity* and *cost* to gain deeper insights into purchasing behaviour. Regarding *Quantity* distribution, there was a notable skew towards lower quantities of items purchased in transactions, indicating that the majority of purchases involved only a few items. This trend highlights the tendency of customers to buy items in small quantities during each transaction. Similarly, the *cost* distribution exhibited a skew towards lower-priced items, suggesting that items with lower prices were more frequently purchased. This observation implies that more affordable items had a higher turnover rate, which is crucial for stocking and recommendation strategies.

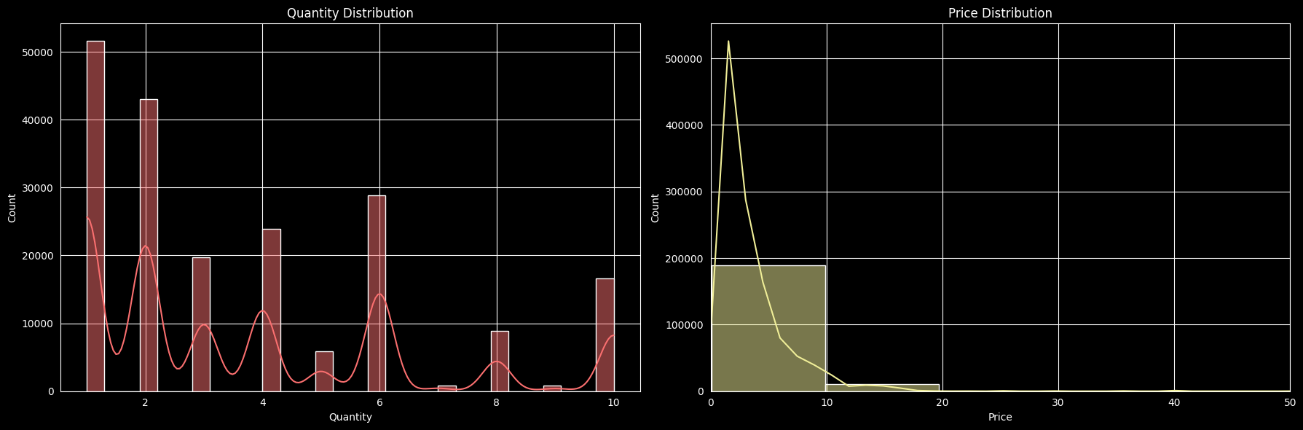


Figure 6 Quantity and price distribution of the dataset

## Top item pairs

Upon reviewing the top item pairs and their purchase frequencies in Table 1, we were able to discern valuable insights into online transaction behaviours, enabling strategic decision-making regarding cross-selling and promotional efforts. Notably, the prevalence of pairings featuring similar items underscores discernible customer preferences and augments our marketing initiatives. Moreover, the abundance of specific categories, illustrated by different types of lunch bags, suggests huge potential for online sales expansion. By leveraging these insights, we can optimise our online sales strategies.

|  |  |  |
| --- | --- | --- |
| **Item 1** | **Item 2** | **Count** |
| LUNCH BAG RED RETROSPOT | LUNCH BAG BLACK SKULL. | 337 |
| LUNCH BAG RED RETROSPOT | LUNCH BAG PINK POLKADOT | 331 |
| LUNCH BAG RED RETROSPOT | LUNCH BAG SPACEBOY DESIGN | 318 |
| ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 311 |
| LUNCH BAG BLACK SKULL. | LUNCH BAG PINK POLKADOT | 303 |
| LUNCH BAG RED RETROSPOT | LUNCH BAG SUKI DESIGN | 297 |
| LUNCH BAG SPACEBOY DESIGN | LUNCH BAG BLACK SKULL. | 287 |
| LUNCH BAG RED RETROSPOT | LUNCH BAG CARS BLUE | 266 |
| SPOTTY BUNTING | PARTY BUNTING | 265 |
| PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 264 |

Table 1 Frequent item pairs and their count

## Analysis of frequently purchased items

Through our analysis, we have identified the most prevalent items in customer transactions, presenting valuable insights into consumer preferences and guiding potential focal points for marketing and recommendation strategies. Our analysis underscored the top 20 frequently purchased items, with the REGENCY CAKESTAND 3 TIER emerging as the leader, closely trailed by WHITE HANGING HEART T-LIGHT HOLDER and PARTY BUNTING. This assortment incorporates a blend of home décor and everyday products, suggesting fields of heightened customer engagement and interest.

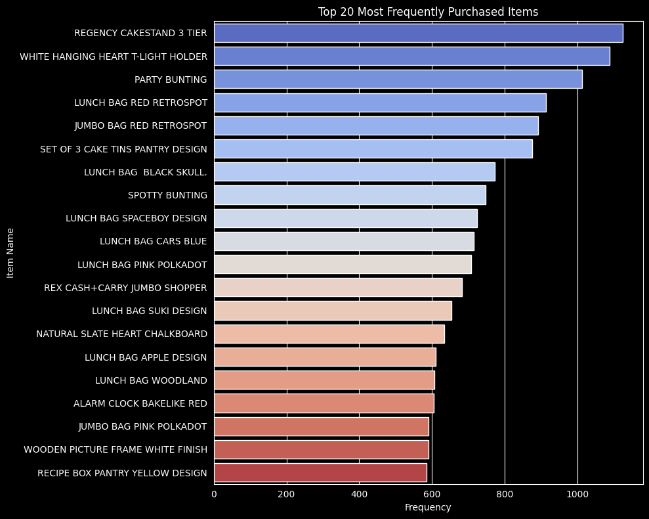


Figure 7 Top 20 most frequently purchased items

## Analysis of temporal trends

Our investigation explored the temporal dynamics within purchasing patterns, aiming to discern potential seasonal or temporal trends related to updating marketing strategies or inventory management decisions. By analysing transaction volumes from December 2010 to October 2011, discernible variations appeared. Peaks in transactional activity coincided with specific months, suggesting probable seasonal influences or promotional impacts on consumer behaviour.

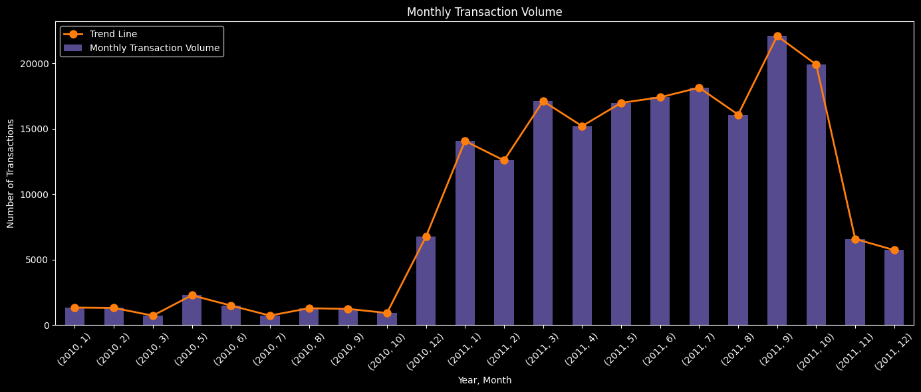


Figure 8 Monthly transaction volume

## Transaction volume by day of the week

The analysis of transaction volumes by day of the week revealed distinct patterns in customer purchasing behaviour. Thursdays and Sundays emerge as the busiest days, with the highest transaction volumes recorded. In contrast, Fridays experience the lowest transaction volume among the days captured in the dataset. Interestingly, Saturdays show no recorded transactions. These insights have significant implications for strategic planning, resource allocation, and marketing initiatives. Leveraging the higher transaction volumes on Thursdays and Sundays for promotional activities or special offers could maximise sales opportunities.

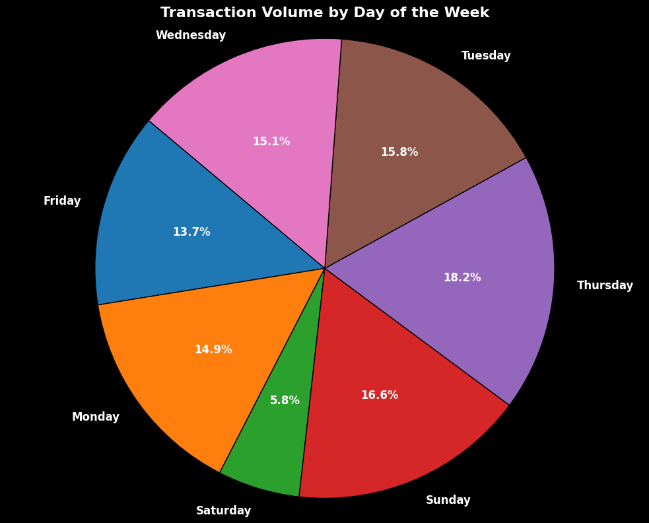


Figure 9 Transaction volume by day of the week

## Top items revenue contribution

To emphasise the importance of high-value items for potential cross-selling, Figure 9 shows the top 10 items contributing the most to revenue in the dataset. It highlights crucial products, such as the REGENCY CAKESTAND 3 TIER, and PARTY BUNTING, which serve as significant sources of revenue and are pivotal to cross-selling and promotional tactics.

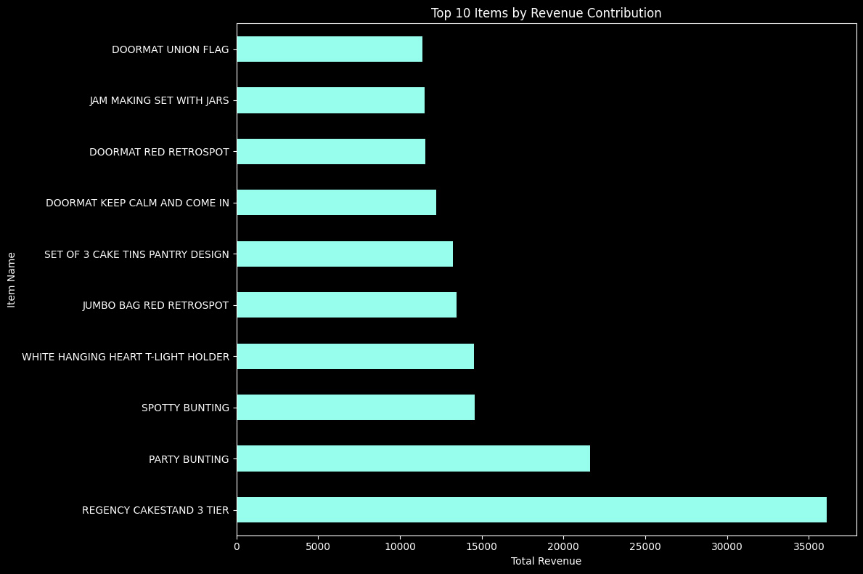


Figure 10 Top 10 items by revenue contribution

# Implementation and Testing

## Recommendation system design

Our recommendation system combines two approaches: collaborative filtering, which looks at item and cosine similarity, and association rule mining, which focuses on frequent item pairs.

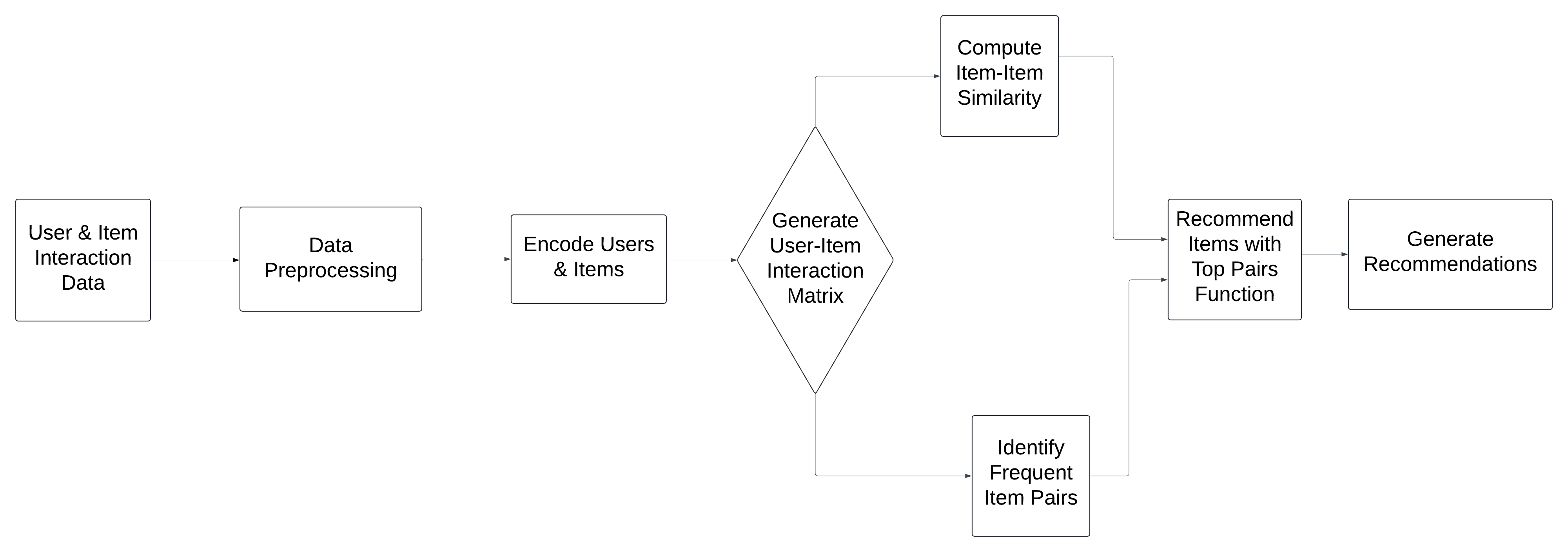


Figure 11 Overview of the recommendation system

The system starts by processing the training data, which involves preparation and organisation, including encoding (see sections [4.2.2](#_Encode_Users_and) and [4.3.1](#_Training_and_Tuning)) *CustomerID* and *Itemname*. Then, it generates a matrix from this pre-processed data, showing how many of each item was bought by each customer, which is crucial for determining item-item similarities.

Using cosine similarity, the system calculates the similarity of each pair of items based on customer purchase behaviour (see section [2.1.1](#_ Cosine_similarity)). Additionally, it looks at which pairs of items are frequently purchased together to improve the accuracy of its recommendations. If there’s an item in the user’s cart that the system doesn’t recognise, it suggests one or more of the most frequently purchased items (Real Python, [2022](#_References)).

In the end, the personalised recommendations are put together for each user to improve their overall experience, cleverly combining collaborative filtering’s ability to understand user preferences based on item similarities and cosine similarity with the insights from association rule mining on item co-occurrences.

## Breakdown of part-wise implementation

## Data preprocessing

Data preprocessing is the primary phase where the user-item interaction data undergoes thorough cleaning and formatting.

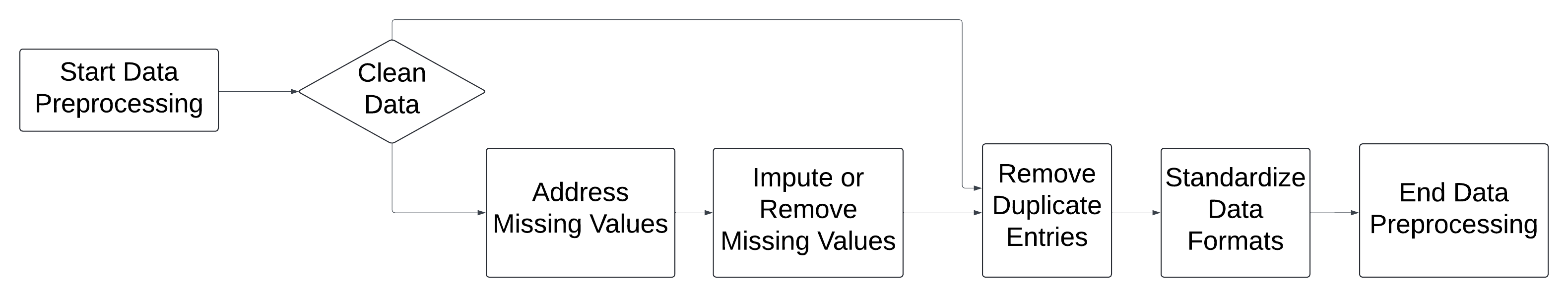


Figure 12 Overview of data preprocessing

This process entails addressing any missing values, removing duplicate entries, and standardising data formats, all aimed at upholding the integrity and quality of the data for the subsequent stages.

## Encoding users and items

Encoding users and items involves converting categorical identifiers into numerical formats suitable for mathematical operations.

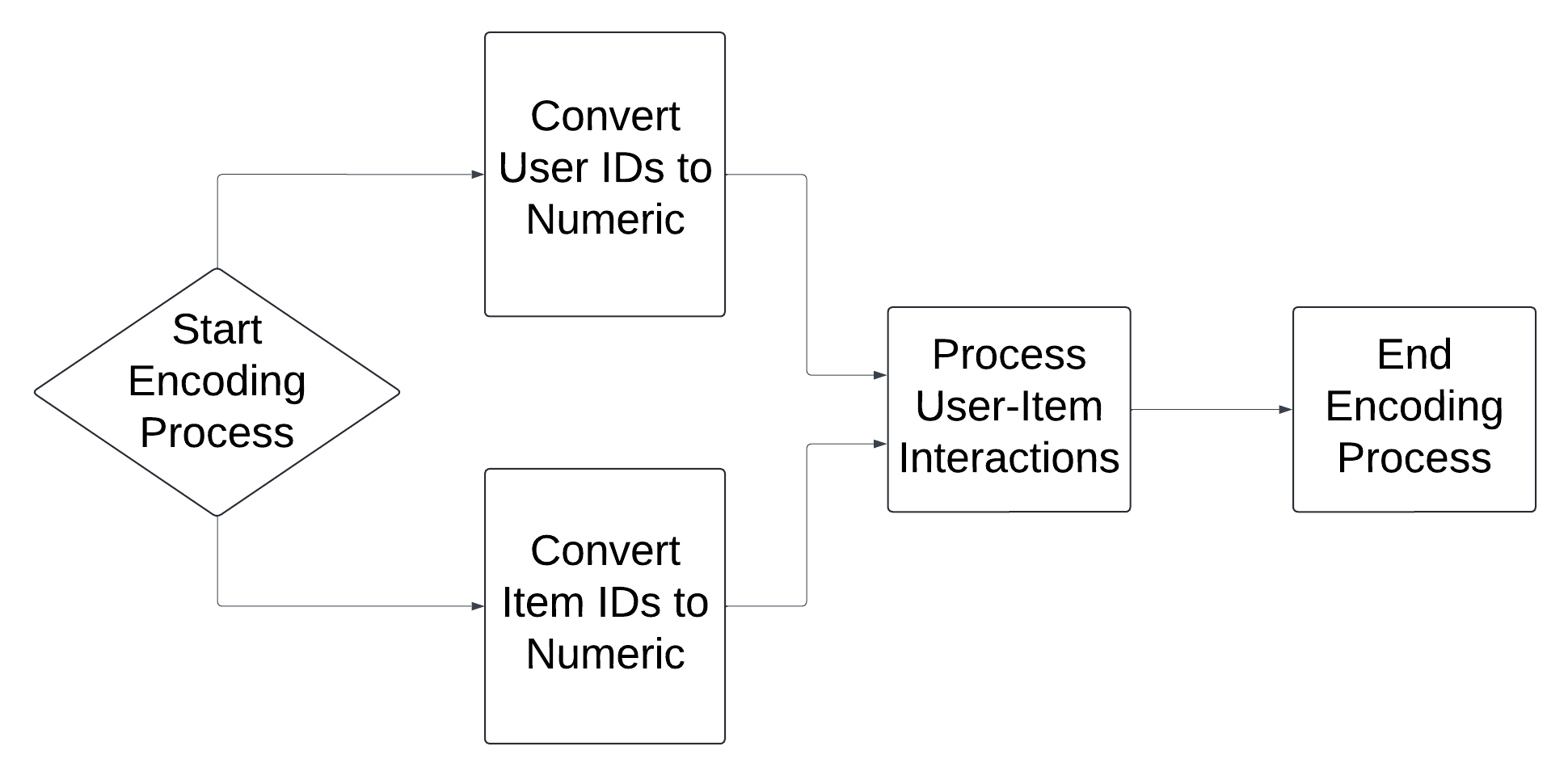


Figure 13 Overview of user and item encoding

This step enables the recommendation system to process and analyse user-item interaction data more efficiently.

## Generating user-item interaction matrix

The user-item interaction matrix acts as a fundamental component, illustrating the interactions between users and items.

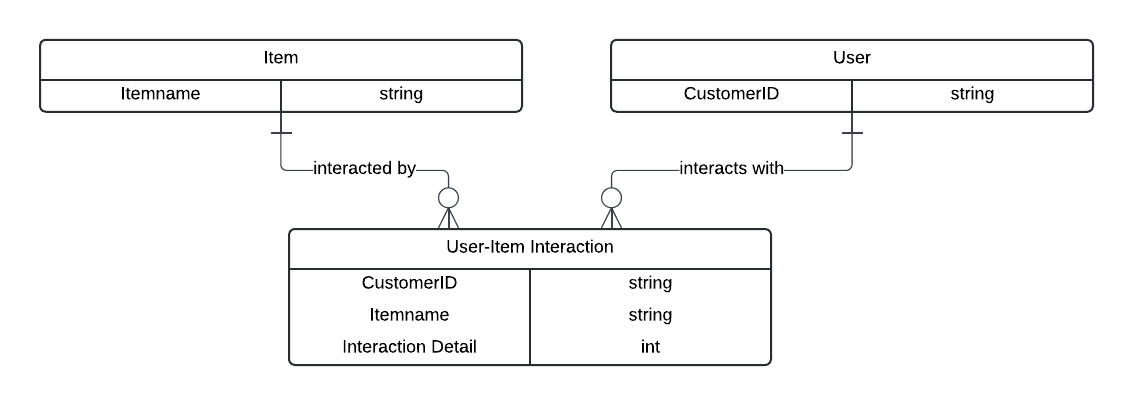


Figure 14 Overview of user-item matrix generation

With users on one axis and items on the other, each cell in the matrix contains essential details about the interaction, such as purchase quantity or rating.

## Computing item-item similarity

Calculating item-item similarity relies on analysing the interactions among items within the user-item interaction matrix.

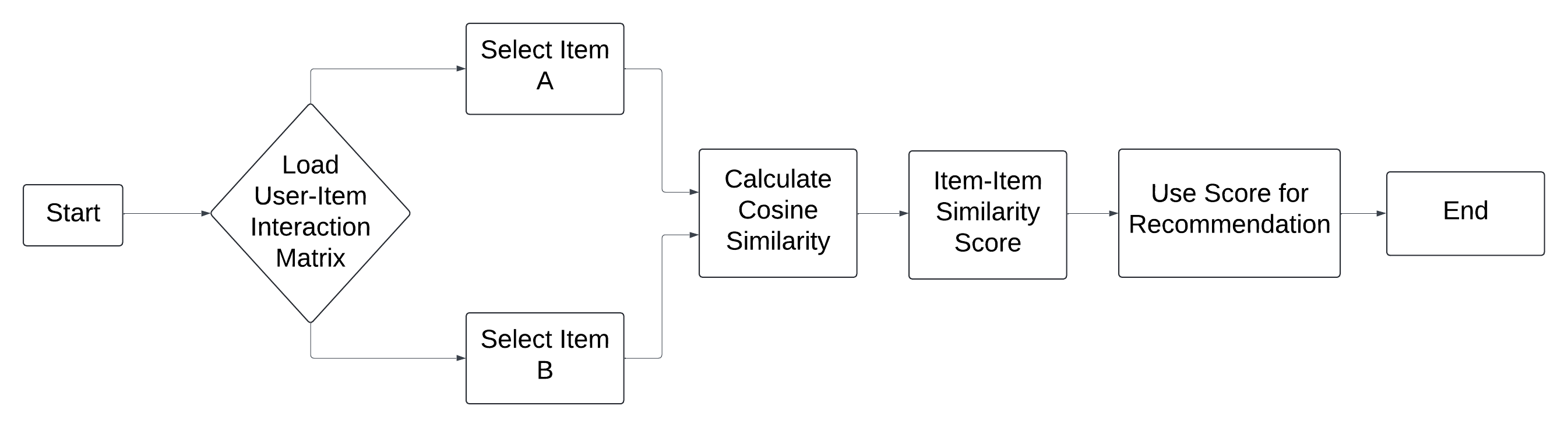


Figure 15 Overview of item-item similarity computation

Commonly used methods, like cosine similarity, are used to measure the resemblance between two items based on user interactions.

## Identifying frequent item pairs

During this stage, frequent item pairs are identified from the user-item interaction data.

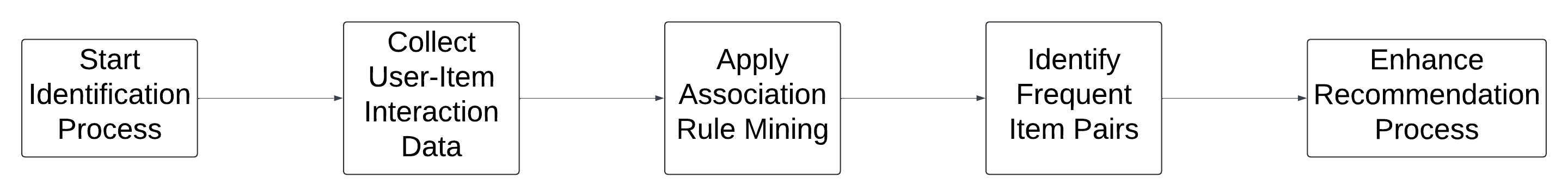


Figure 16 Overview of identification of frequent item pairs

By leveraging methods like association rule mining, these frequent item pairs signify strong correlations among items, indicating important associations between them.

## Personalising recommendations

The recommendation function integrates item-item similarity and frequent item pairs to produce personalised recommendations for items in the user’s cart. Using this, the system provides diverse and pertinent recommendations while reducing redundancy.

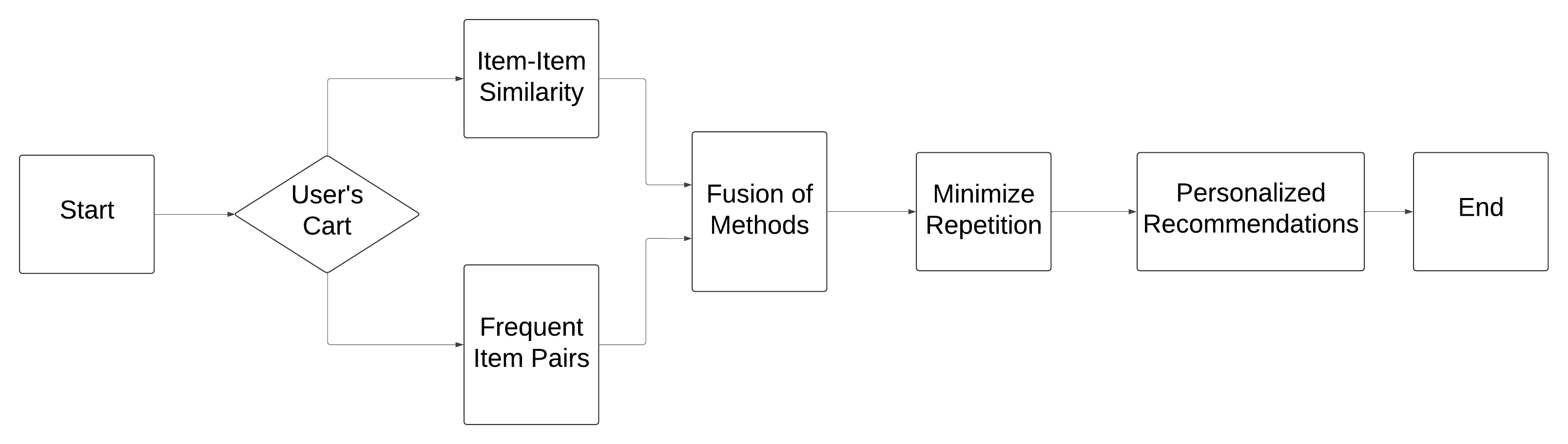


Figure 17 Overview of personalised recommendation generation

## Generating recommendations

After the recommendation function is executed, the system generates recommendations based on the combined results.

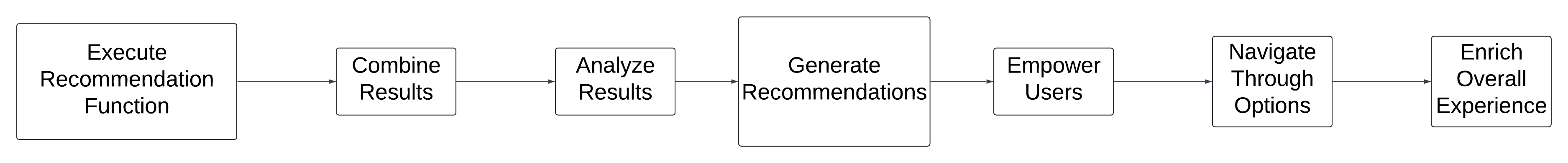


Figure 18 Overview of recommendation generation

Upon thorough analysis of these recommendations, the users are efficiently able to navigate through a wide range of options, thereby enriching their overall experience.

## Methodology

## Training and tuning on the training set

The training set, consisting of user-item interactions over a specific period, was used to construct and fine-tune our recommendation system. This dataset was used to:

* **Encode customer and item data:** Both *CustomerID* and *Itemname* were encoded using label encoding to facilitate numerical processing.
* **Create a user-item matrix:** A sparse matrix representing the quantity of each item purchased by each customer was constructed to serve as the foundation for calculating item similarities and generating recommendations.
* **Calculate item-item similarity:** Cosine similarity was used to compute a similarity matrix among items based on their co-purchase patterns by different customers.

## Using the test set for final results

The test set, which comprised of a separate set of user-item interactions, was deployed to evaluate the recommendation system’s performance. This dataset allowed us to assess the system’s ability to generate relevant recommendations under conditions that simulated real-world usage.

## Evaluation methodology

The evaluation focused on determining the system’s effectiveness in suggesting items similar to those in a user’s cart, but not currently included. The primary steps involved were:

* **Sampling users:** To ensure focused evaluation, we selected user carts with 5-7 items, balancing simplicity and complexity.
* **Masking strategy:** We masked an item in each cart based on frequent co-purchased pairs and its average similarity to other items, prioritising those most representative of the user’s selection.
* **Generating recommendations:** Recommendations for the remaining items were based on their similarity to the masked item and frequent item pairs.
* **Evaluation:** Success was measured by recommending items with a similarity score of 0.5 or higher to the masked item, quantitatively assessing recommendation relevance.

## Discussion of Results

## Frequent patterns

## Training set patterns

|  |  |
| --- | --- |
| **Item Pair** | **Support** |
| LUNCH BAG BLACK SKULL., LUNCH BAG RED RETROSPOT | 0.73703 |
| LUNCH BAG RED RETROSPOT, LUNCH BAG SUKI DESIGN | 0.680758 |
| LUNCH BAG CARS BLUE, LUNCH BAG RED RETROSPOT | 0.656053 |
| LUNCH BAG APPLE DESIGN, LUNCH BAG RED RETROSPOT | 0.653857 |
| LUNCH BAG PINK POLKADOT, LUNCH BAG RED RETROSPOT | 0.643151 |

Table 2 Top 5 frequent item pairs in the train dataset and their support

The dominance of various LUNCH BAG combinations in the training set patterns reveals significant cross-selling opportunities, and highlights the importance of inventory diversity to cater to varied customer preferences. Monitoring these patterns can help in effective inventory management and highlight emerging trends, enabling proactive adjustments to product offerings and marketing strategies. This approach not only improves the shopping experience by making it more intuitive but also supports increased average order values through informed product bundling and promotions.

## Testing set patterns

Our analysis identified the top 5 frequent item pairs in the train dataset primarily around various LUNCH BAG designs, which contrasts with the varied patterns implied through recommendations in the test set data (see Table 4). This insight is vital, as it demonstrates the variety and distinct purchasing habits within our customer base. In real-world scenarios, these patterns might include different item combinations that vary seasonally or due to shifting consumer preferences.

## 10 examples of recommendations

Based on the frequent item pairs identified, the recommendation system could suggest complementary items when one item in a pair is present in a user’s cart. Considering 3 hypothetical scenarios:

* If user has LUNCH BAG BLACK SKULL in their cart, recommend LUNCH BAG RED RETROSPOT as they are frequently purchased together.
* If user has LUNCH BAG RED RETROSPOT in their cart, recommend LUNCH BAG SUKI DESIGN due to high co-purchase frequency.
* If user has LUNCH BAG CARS BLUE in their cart, recommend LUNCH BAG RED RETROSPOT as they are frequently purchased together.

Using this logic, the recommendation system repeats this for each of the top 5 pairs, tailoring recommendations based on what’s already in the user’s cart.

|  |  |
| --- | --- |
| **Item Pair** | **Support** |
| LUNCH BAG BLACK SKULL., LUNCH BAG RED RETROSPOT | 0.73703 |
| LUNCH BAG RED RETROSPOT, LUNCH BAG SUKI DESIGN | 0.680758 |
| LUNCH BAG CARS BLUE, LUNCH BAG RED RETROSPOT | 0.656053 |
| LUNCH BAG APPLE DESIGN, LUNCH BAG RED RETROSPOT | 0.653857 |
| LUNCH BAG PINK POLKADOT, LUNCH BAG RED RETROSPOT | 0.643151 |
| LUNCH BAG BLACK SKULL., LUNCH BAG CARS BLUE | 0.64123 |
| LUNCH BAG RED RETROSPOT, LUNCH BAG SPACEBOY DESIGN | 0.631348 |
| CHARLOTTE BAG SUKI DESIGN, RED RETROSPOT CHARLOTTE BAG | 0.619819 |
| JUMBO BAG RED RETROSPOT, LUNCH BAG RED RETROSPOT | 0.601427 |
| LUNCH BAG BLACK SKULL., LUNCH BAG SUKI DESIGN | 0.597584 |

Table 3 Top 10 frequent patterns in the training set and their support

|  |  |
| --- | --- |
| **Item Pair** | **Support** |
| GARDENERS KNEELING PAD CUP OF TEA, SMALL HANGING IVORY/RED WOOD BIRD | 0.212427 |
| GARDENERS KNEELING PAD CUP OF TEA, GARDENERS KNEELING PAD KEEP CALM | 0.204992 |
| LIPSTICK PEN RED, SMALL HANGING IVORY/RED WOOD BIRD | 0.19384 |
| PAPER CHAIN KIT 50’S CHRISTMAS, PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.190122 |
| CHOCOLATE HOT WATER BOTTLE, HOT WATER BOTTLE KEEP CALM | 0.181625 |
| GARDENERS KNEELING PAD CUP OF TEA, HOT WATER BOTTLE KEEP CALM | 0.177908 |
| HOT WATER BOTTLE KEEP CALM, SMALL HANGING IVORY/RED WOOD BIRD | 0.17419 |
| GARDENERS KNEELING PAD CUP OF TEA, LIPSTICK PEN RED | 0.172597 |
| SET OF 3 WOODEN STOCKING DECORATION, SMALL HANGING IVORY/RED WOOD BIRD | 0.169942 |
| SKULL DESIGN TV DINNER TRAY, SMALL HANGING IVORY/RED WOOD BIRD | 0.167286 |

Table 4 Top 10 frequent patterns in the testing set and their support

From the analysis done on the train and test datasets with pattern mining, we have observed some key insights:

* For SET OF 6 SPICE TINS PANTRY DESIGN, the system recommended items like RETROSPOT CHILDRENS APRON and CHILDREN’S APRON DOLLY GIRL with similarity scores indicating a decent match based on pattern mining.
* For FAMILY ALBUM WHITE PICTURE FRAME, suggestions included WALL ART SPACEBOY and WALL ART BUFFALO BILL, displaying how pattern mining can also give non-obvious recommendations beyond direct item matches.

These recommendations show the system’s capacity to leverage identified patterns for suggesting items that, while not always directly related, might still hold interest for the shopper based on the observed buying behaviours of others.

## Discussion of metrics

## Performance with and without pattern mining

To test the efficiency of our recommendation system, we also created a control method that gave recommendations without the use of any pattern mining algorithms. Comparing the results of both of these methods, we were able to discern the following:

* When utilising pattern mining, the recommendation system achieved a performance score of 0.50, indicating its effectiveness in delivering relevant recommendations to users. The system was able to make informed suggestions that closely aligned with user preferences.
* In contrast, the recommendation system without pattern mining achieved a lower performance score of 0.40, suggesting that the absence of pattern mining resulted in less effective recommendations.

This analysis highlights the importance of pattern mining in recommendation systems and emphasises its role in enhancing recommendation relevance and user experience (Evidently AI, [2024](#_References)).

## Impact of pattern mining

The performance score of 0.50 using pattern mining highlights its pivotal role in reinforcing recommendation relevance and empowers the system to glean valuable insights from user behaviour, uncovering meaningful associations among items frequently purchased together. The disparity in performance scores highlights the significant role of pattern mining in improving recommendation system performance. By identifying and leveraging item associations, pattern mining enables the system to offer more tailored and personalised suggestions, ultimately enhancing user satisfaction and engagement.

## Conclusion and Recommendations

Throughout this report, we highlighted the benefits of integrating our hybrid recommendation system, which combined pattern mining with collaborative filtering and provided significant benefits, to online grocery retailer. Our method allowed for personalised product recommendations by analysing customer purchasing habits and preferences. This approach not only serves to improve user experience by making shopping more intuitive and satisfying, but also offers potential benefits in terms of increased sales and customer loyalty for e-commerce platforms. By tailoring suggestions based on popular item pairs, our system not only increased user satisfaction, but also encouraged repeated purchases and client loyalty, as well as introduced new customers to the store’s best-selling items (Holewa, [2023](#_References)). Overall, our hybrid recommendation system provided excellent potential for online grocery retailer to increase customer engagement and drive business growth.

## Addressing challenges

While association rule mining has numerous benefits, it also has drawbacks, such as the potential to create an overwhelming number of rules from huge datasets. To address this issue, techniques such as establishing minimum support and confidence levels, using hierarchical structures, and implementing pruning and ranking procedures can help minimise the search area and eliminate irrelevant or redundant rules.

In addition, association rule mining is dependent on parameter selection. To overcome this, domain and context, data qualities and distribution, and evaluation and validation methodologies must all be considered to provide accurate and useful results (LinkedIn, [2023a](#_References)). Finally, association rule mining might generate a vast number of rules, which can be difficult to analyse and understand. Techniques including visualisation, grouping or classification, and narration can help analyse and communicate results more effectively, allowing for better insights from the data.

## Recommendations for improvement

Further refining of the pattern mining process could involve exploring more nuanced patterns, or considering temporal factors (e.g., seasonal variations in item pair popularity). Additionally, integrating user feedback mechanisms could help in dynamically adjusting the recommendations to better align with evolving user preferences. Analysing the similarity scores of recommended items can provide additional insights into the strength of associations and the effectiveness of recommendations, guiding further optimisations and refinements to the recommendation algorithm.

In future development, one area where we could improve is in streamlining the codebase and optimising the efficiency of the recommendation algorithms. While the current implementation addresses the core functionalities effectively, there may be better algorithms to enhance scalability and performance, especially when dealing with larger datasets or deploying the system in real-time environments. Employing advanced optimisation techniques, leveraging distributed computing frameworks, or exploring alternative algorithms could potentially enhance the system’s speed and robustness.

## Reflection

Through this project, we have learned the importance of addressing various challenges present in recommendation systems. Our previous attempts at building recommendation systems ranged from simplistic approaches like suggesting the same items to all users to random item selection, none of which yielded satisfactory results. From encoding categorical variables to computing similarity matrices and identifying frequent item pairs, each step contributed to improving recommendation accuracy and relevance. One main takeaway was the significance of leveraging various techniques, such as collaborative filtering and association rule mining, to tackle issues like the cold start problem, data sparsity, and recommendation diversity. The incorporation of top frequent item pairs into the recommendation process enhanced the system’s ability to suggest complementary items, thereby improving user satisfaction and engagement.

This experience has been immensely rewarding; the complexity of the task pushed us to explore new horizons and think innovatively. By applying these algorithms in a practical setting, we not only deepened our understanding of them, but also honed our problem-solving skills.

# References

Afsar, M.M., Crump, T. and Far, B. (2022) ‘Reinforcement learning based Recommender Systems: A survey’, *ACM Computing Surveys*, 55(7), pp. 1–38. doi:10.1145/3543846.

AI & LinkedIn Community (2023a) *How can you overcome the challenges of association rule mining?*, *LinkedIn*. Available at: https://www.linkedin.com/advice/0/how-can-you-overcome-challenges-association-rule-mining (Accessed: 26 March 2024).

AI & LinkedIn Community (2023b) *How do you use association rule mining for cross-selling and recommendation systems?*, *LinkedIn*. Available at: https://www.linkedin.com/advice/0/how-do-you-use-association-rule-mining (Accessed: 26 March 2024).

Evidently AI Team (2024) *10 metrics to evaluate recommender and ranking systems*, *Evidently AI - Open-Source ML Monitoring and Observability*. Available at: https://www.evidentlyai.com/ranking-metrics/evaluating-recommender-systems (Accessed: 02 April 2024).

Gambito Analytics (2023) *Unleashing the power of personalization: Advantages of recommendation systems for diverse businesses*, *LinkedIn*. Available at: https://www.linkedin.com/pulse/unleashing-power-personalization-advantages-recommendation/ (Accessed: 27 March 2024).

GeeksforGeeks (2023) *Classification using frequent patterns in data mining*, *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/classification-using-frequent-patterns-in-data-mining/ (Accessed: 04 April 2024).

Herbold, S. (2020) ‘Association Rule Mining’, in *Introduction to Data Science Release 0.1*. Steffen Herbold, pp. 55–69. Available at: https://sherbold.github.io/intro-to-data-science/pdf/introductiontodatascience.pdf (Accessed: 05 April 2024).

Holewa, K. (2023) *Recommendation systems: Benefits, Types & Examples* , *Miquido*. Available at: https://www.miquido.com/blog/perks-of-recommendation-systems-in-business/ (Accessed: 24 March 2024).

Kirk, J. (2019) *Getting started with recommender systems and TensorRec*, *Medium*. Available at: https://towardsdatascience.com/getting-started-with-recommender-systems-and-tensorrec-8f50a9943eef (Accessed: 06 April 2024).

Kryzwicki, A. (2024) [Lecture] *COMP\_SCI\_7306: Mining Big Data*. The University of Adelaide. March 2024.

Prabhakaran, S. (2022) *Cosine similarity - understanding the math and how it works? (with python)*, *machine learning +*. Available at: https://www.machinelearningplus.com/nlp/cosine-similarity/ (Accessed: 01 April 2024).

Real Python (2022) *Build a recommendation engine with collaborative filtering*, *Real Python*. Available at: https://realpython.com/build-recommendation-engine-collaborative-filtering (Accessed: 01 April 2024).

Silva, X. dos S. (2020) *Recommender Systems for Grocery Retail [A Machine Learning Approach]*, *Instituto Superior de Engenharia do Porto (ISEP) - Instituto Politécnico do Porto (IPP)*. Available at: https://recipp.ipp.pt/bitstream/10400.22/17716/1/DM\_XavierSilva\_2020\_MEI.pdf (Accessed: 24 March 2024).

Verma, P. and Sharma, S. (2021) ‘Artificial Intelligence based recommendation system’, *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 669–673. doi:10.1109/icacccn51052.2020.9362962.

Zhang, Q., Lu, J. and Jin, Y. (2020) ‘Artificial Intelligence in Recommender Systems’, *Complex & Intelligent Systems*, 7(1), pp. 439–457. doi:10.1007/s40747-020-00212-w.

# *Appendix 1*

## Glossary of Terms

**Association rule mining:** A technique for discovering interesting relationships between variables in large databases. For example, people who buy bread often also buy butter.

**Big data:** Vast amounts of data that are too complex for traditional data processing tools to handle. Big data can reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

**Cold start problem:** The difficulty recommendation systems have in making accurate suggestions when there’s not enough data on new users or items.

**Collaborative filtering:** A method used by recommendation systems to suggest items by analysing the preferences or behaviours of many users. If User A likes the same items as User B, then the system might recommend to User A an item that User B likes.

**Cosine similarity:** A measure used to calculate the similarity between two items or users, considering their preferences or behaviours as vectors in a multi-dimensional space. It’s like measuring how close two arrows are to pointing in the same direction.

**Data preprocessing:** The process of cleaning and organising raw data into a suitable format for analysis. This might include removing errors, filling in missing values, or converting data into a standard format.

**Encoding:** Transforming categorical data (like item names) into numerical form so that it can be processed by algorithms.

**Frequent itemsets:** Groups of items that appear together in a dataset more often than expected by chance. This concept is often used in market basket analysis to find items that are commonly purchased together.

**Pattern mining:** The process of finding repeated patterns within data, such as frequently bought together items in a shopping dataset.

**Recommendation system:** A type of information filtering system that suggests items (like movies, books, products) to users based on various criteria, such as past behaviour, similarities with other users, or item features.

**Scalability:** The capability of a system to handle a growing amount of work by adding resources to the system. In the context of recommendation systems, it refers to the system’s ability to efficiently process increasing amounts of data or serve a growing number of users.

**User-item interaction matrix:** A grid that represents the interactions between users and items, such as purchase history or ratings, where rows represent users and columns represent items.